



Title:

Application of Machine Learning Algorithms for the Detection of Spatio-Temporal Patterns linked to Geophysical Phenomena in GNSS Time Series

Author(s): *Laura Crocetti*

Name of Academic Institution: ETH Zurich

Level of study or work: PhD thesis

(Bachelor thesis, master, PhD, research, project, etc.)

Information about you (and your team): My name is Laura Crocetti, and I am currently a postdoctoral researcher at ETH Zurich in the Space Geodesy group of Prof. Soja at the Institute of Geodesy and Photogrammetry. In December 2024, I successfully finished my PhD in the same group. My doctoral research explored the novel integration of machine learning with Global Navigation Satellite Systems (GNSS) and environmental data. It featured innovative, cross-disciplinary approaches that contributed to new insights in both geospatial analysis and environmental monitoring. Before starting my PhD, I worked as a Project Assistant at TU Wien in the Climate and Environmental Remote Sensing research group at the Department of Geodesy and Geoinformation, where I also completed my bachelor's and master's degrees in Geodesy and Geoinformation.

Area of interest

(Identifying the problem, explain why it is important and the current relevance of the topic, up to 250 words)

Global Navigation Satellite Systems (GNSS) are best known for positioning, navigation, and timing, but GNSS observations also contain valuable geophysical information that can be extracted and used to improve our understanding of the Earth system. A huge amount of data is available with tens of thousands of permanent GNSS stations distributed globally. To fully exploit the wealth of the available GNSS data, artificial intelligence (AI) and, in particular, machine learning (ML) offer a promising opportunity. Although AI and ML are already well established in many scientific disciplines, they have only recently been explored in space geodesy.

The thesis [1] covers three major research areas, with a focus on the development of a global ML model for the spatio-temporal prediction of GNSS zenith wet delay (ZWD). ZWD is a critical parameter that represents one of the biggest error sources for space geodetic techniques and is strongly correlated with the water vapour content in the atmosphere. Thus, accurate modelling of ZWD is of particular interest to improve positioning and navigation, weather forecasting, and for climate change studies, as it provides detailed insights into the Earth's atmospheric conditions. A disadvantage, however, is that ZWD measurements are allocated point-wise and can only be retrieved at the GNSS station locations, whose spatial distribution is very irregular. This work aims to overcome this limitation. The main objective is to use the available datasets and derive a continuous and globally available ZWD model that fully captures the Earth's system, using novel ML approaches.



Approach to the problem

(Describe your methodology or technology and how it will solve the problem you identified, up to 300 words)

To build a global ZWD model, we formulate a regression problem based on the XGBoost algorithm. It is an efficient, scalable ML algorithm featuring an ensemble of regression trees. These trees are iteratively fitted to the dataset using a technique called gradient boosting. The choice of algorithm is driven by extensive empirical investigations. The best hyperparameters are found by utilizing a systematic approach that exhaustively investigates parameter combinations and evaluates their performance.

The developed model, called ZWDX, is trained on 12 years of data (2010-2021) using hourly ZWD estimates of over 19.000 globally distributed GNSS stations provided by the Nevada Geodetic Laboratory. The input features for the model are the GNSS station locations (latitude, longitude, height), relative time information (day-of-year, hour-of-day), and hourly specific humidity data on nine pressure levels, ranging from 1000 to 500 hPa. The specific humidities are provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) on a regular latitude-longitude grid with a spatial resolution of 0.25 degrees, which are bilinearly interpolated to the GNSS station locations. To account for the cyclic nature of the longitude, day-of-year, and hour-of-day, these features are transformed into pairs of $\sin(\cdot)$ and $\cos(\cdot)$ values. This leads to a total of seventeen features, which are all standardized before feeding them into the XGBoost model.

The GNSS stations are randomly split into 16.664 training stations and 2.500 test stations, resulting in ~ 820 million training samples and ~ 247 million test samples. To our knowledge, this makes it one of the most comprehensive ZWD datasets used to date for this type of analysis. The geographic location of the GNSS stations is depicted in Figure 1.

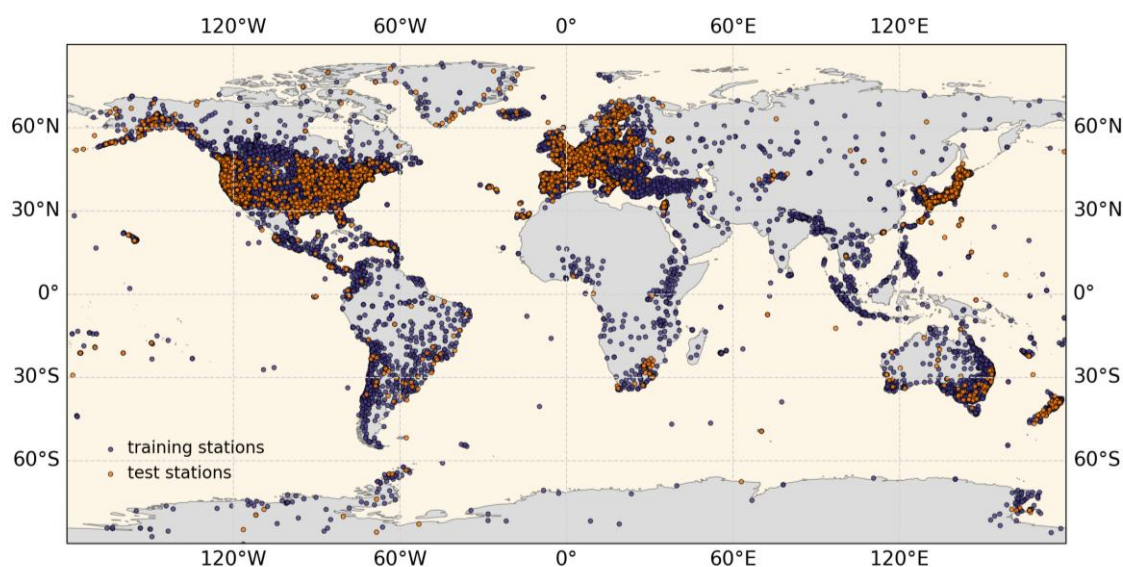


Figure 1: Distribution of the used GNSS stations, divided into training (blue) and test (orange) stations

Figure 2 illustrates the workflow of the ML setup. More details on the method can be found in the corresponding papers [2, 3].

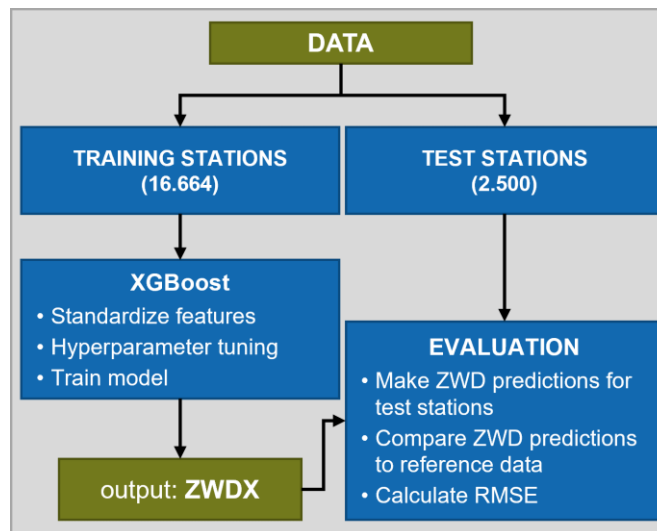


Figure 2: ML setup. The GNSS data is split into training and test stations. Training stations are used to train the ZWDX model. Test stations are used for evaluation.

Results, conclusions and next steps

(Present your results and conclusions of your study, up to 250 words)

ZWDX achieves a root mean squared error of 10.8 mm on the test stations for the year 2022. Note that this year is not included in the training process, ensuring temporal and spatial independence.

The ZWDX model can be applied at any desired location on Earth and for any point in time if the input variables are available. As an example, we have predicted a global ZWD map on a 0.25 degrees latitude/longitude grid for July 1, 2022, at 00:00 UTC, see Figure 3. The ZWD map reveals the expected large-scale patterns, with overall higher values in the tropics and lower values in the polar regions, attributed to the higher/lower water vapour in the atmosphere.

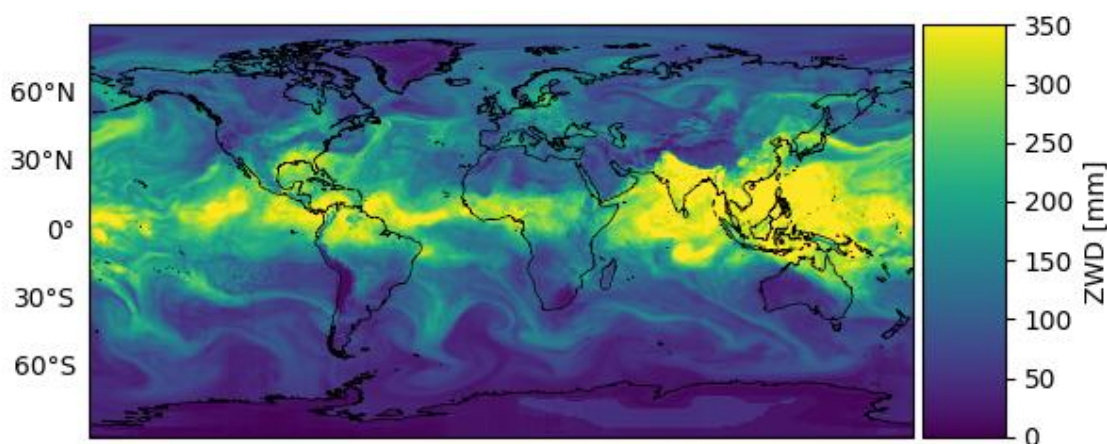


Figure 3: Global ZWD predictions for July 1, 2022, 00:00 UTC

To demonstrate the value of ZWDX, the model has been used in GNSS Precise Point Positioning (PPP), where it achieved slightly higher accuracies than state-of-the-art models.

Furthermore, ZWDX's ability to efficiently generate global ZWD fields and time series is valuable to neighboring research domains, such as atmosphere and Earth system science. In a complementary study [4], we used the model to compute global ZWD trends over time for climate studies and could confirm a positive trend in the Northern Hemisphere and a negative trend in the Southern Hemisphere, see Figure 4.

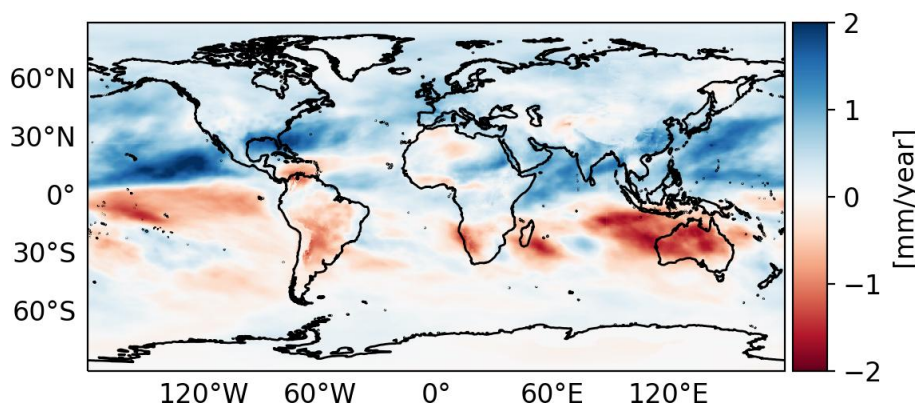


Figure 4: Global ZWD trend per grid point from 2010 to 2023

In summary, ZWDX is a global ML model enabling spatio-temporal prediction of ZWD. It has demonstrated potential for GNSS processing and for global climate analyses. The ZWDX model and its predictions are publicly available at the Geodetic Prediction Center of ETH Zurich.

References

(Additional information, publications, or links, up to 200 words, optional)

[1] Crocetti, L. Application of Machine Learning Algorithms for the Detection of Spatio-Temporal Patterns linked to Geophysical Phenomena in GNSS Time Series. [ETH Zurich] (2024). <https://doi.org/10.3929/ETHZ-B-000715655>

[2] Crocetti, L., Schartner, M., Wareyka-Glaner, M.F. et al. ZWDX: a global zenith wet delay forecasting model using XGBoost. *Earth Planets Space* **76**, 163 (2024). <https://doi.org/10.1186/s40623-024-02104-6>

[3] Crocetti, L., Schartner, M., Zus, F. et al. Global, spatially explicit modelling of zenith wet delay with XGBoost. *J Geod* **98**, 23 (2024). <https://doi.org/10.1007/s00190-024-01829-2>

[4] Crocetti, L., Schartner, M., Schindler, K. et al. Modelling the Troposphere with Global Navigation Satellite Systems, Meteorological Data and Machine Learning. *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 1689-1692 (2024). <https://doi.org/10.1109/IGARSS53475.2024.10640441>

END